Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

Claims 1-23 (Cancelled)

24. (New) A method for retrieving information using a search engine, the method comprising:

retrieving a document to be indexed;

generating a virtual document based on the retrieved document, the virtual document comprising a portion of the retrieved document that characterizes an overall content of the retrieved document and being used to index the retrieved document;

decomposing the virtual document into a plurality of tokens; and

storing the plurality of tokens in a search index, wherein the search engine accesses the search index to identify one or more virtual documents that satisfy a search query and retrieves one or more documents corresponding to the one or more virtual documents.

25. (New) The method of claim 24, further comprising:

recording position information relating to the portion of the retrieved document that characterizes the overall content of the retrieved document.

26. (New) The method of claim 25, further comprising:

storing the recorded positional information with the plurality of tokens in the search index.

27. (New) The method of claim 24, wherein the portion of the retrieved document that characterizes the overall content of the retrieved document is a summary of the retrieved document.

28. (New) The method of claim 24, wherein generating the virtual document based on the retrieved document comprises:

extracting from the retrieved document a collection of words, features, whole sentences, or parts of sentences that characterizes the overall content of the retrieved document.

29. (New) The method of claim 28, wherein extraction of the collection of words, features, whole sentences, or parts of sentences is based on frequency of occurrence, proximity to the beginning or end of a paragraph, proximity to the beginning or end of the retrieved document, or position within a certain document structure in the retrieved document.

30. (New) The method of claim 24, wherein each of the plurality of tokens comprises a word, a feature, a whole sentence, or a part of a sentence in the virtual document.

31. (New) The method of claim 24, wherein the retrieved document is a web-page.

32. (New) A computer readable medium containing a computer program for retrieving information using a search engine, the computer program comprising program instructions for:

retrieving a document to be indexed;

generating a virtual document based on the retrieved document, the virtual document comprising a portion of the retrieved document that characterizes an overall content of the retrieved document and being used to index the retrieved document;

decomposing the virtual document into a plurality of tokens; and

storing the plurality of tokens in a search index, wherein the search engine accesses the search index to identify one or more virtual documents that satisfy a search query and retrieves one or more documents corresponding to the one or more virtual documents.

33. (New) The computer readable medium of claim 32, wherein the computer program further comprises program instructions for:

recording position information relating to the portion of the retrieved document that characterizes the overall content of the retrieved document.

34. (New) The computer readable medium of claim 33, wherein the computer program further comprises program instructions for:

storing the recorded positional information with the plurality of tokens in the search index.

35. (New) The computer readable medium of claim 32, wherein the portion of the retrieved document that characterizes the overall content of the retrieved document is a summary of the retrieved document.

36. (New) The computer readable medium of claim 32, wherein generating the virtual document based on the retrieved document comprises:

extracting from the retrieved document a collection of words, features, whole sentences, or parts of sentences that characterizes the overall content of the retrieved document.

37. (New) The computer readable medium of claim 36, wherein extraction of the collection of words, features, whole sentences, or parts of sentences is based on frequency of occurrence, proximity to the beginning or end of a paragraph, proximity to the beginning or end of the retrieved document, or position within a certain document structure in the retrieved document.

38. (New) The computer readable medium of claim 32, wherein each of the plurality of tokens comprises a word, a feature, a whole sentence, or a part of a sentence in the virtual document.

39. (New) The computer readable medium of claim 32, wherein the retrieved document is a web-page.

40. (New) A system for retrieving information using a search engine, the system comprising:

a crawler for retrieving a document to be indexed;

an extractor coupled to the crawler for generating a virtual document based on the retrieved document, the virtual document comprising a portion of the retrieved document that characterizes an overall content of the retrieved document and being used to index the retrieved document;

a storage device coupled to the extractor for storing the virtual document;

an indexer coupled to the storage device for decomposing the virtual document into a plurality of tokens; and

a search index coupled to the indexer for storing the plurality of tokens, wherein the search engine accesses the search index to identify one or more virtual documents that satisfy a search query and retrieves one or more documents corresponding to the one or more virtual documents.

41. (New) The system of claim 40, wherein the extractor records position information relating to the portion of the retrieved document that characterizes the overall content of the retrieved document.

42. (New) The system of claim 41, wherein the search index stores the recorded positional information with the plurality of tokens.

43. (New) The system of claim 40, wherein the portion of the retrieved document that characterizes the overall content of the retrieved document is a summary of the retrieved document.

44. (New) The system of claim 40, wherein generating the virtual document based on the retrieved document comprises:

extracting from the retrieved document a collection of words, features, whole sentences, or parts of sentences that characterizes the overall content of the retrieved document.

45. (New) The system of claim 44, wherein extraction of the collection of words, features, whole sentences, or parts of sentences is based on frequency of occurrence, proximity to the beginning or end of a paragraph, proximity to the beginning or end of the retrieved document, or position within a certain document structure in the retrieved document.

46. (New) The system of claim 40, wherein each of the plurality of tokens comprises a word, a feature, a whole sentence, or a part of a sentence in the virtual document.

47. (New) The system of claim 40, wherein the retrieved document is a web-page.